УДК 681.513

РОЗРОБКА ТА ДОСЛІДЖЕННЯ МЕТОДУ ІНДУКТИВНОГО МОДЕЛЮВАННЯ З НЕЧІТКИМИ ВХІДНИМИ ЗМІННИМИ

Д.А. ПІНЧУК

Представлено метод групового урахування аргументів з нечіткими входами. Запропоновано розв'язання задачі прогнозування доходу універсаму. Показано порівняльні результати з нечітким методом групового урахування аргументів.

ВСТУП

Підчас прийняття рішень про інвестування великого проекту з довгим періодом окупності дуже важливо вірно оцінити економічну ефективність та спрогнозувати очікуваний результат.

Саме тому активно розвиваються методи і методики моделювання та прогнозування. Методи відрізняються за типом та об'ємом вхідних даних, внутрішнім принципом організації.

Наша економіка з тих чи інших причин дуже швидко набуває змін. Через це для побудови моделей та прогнозування в умовах економіки України доцільно використовувати ті методи, що базуються на спостереженнях протягом невеликого часу. Одним із вдалих у цьому плані є метод групового урахування аргументів (МГУА) — метод самоорганізації, викладений О.Г. Івахненком у роботах [1,2].

Дуже цікавою модифікацією МГУА є нечіткий МГУА (НМГУА), розроблений в Інституті прикладного системного аналізу Ю.П. Зайченко [3]. Ця модифікація вдосконалює метод. Вона відкидає обмеження на вродженість вхідної матриці даних на відміну від МГУА.

Обидві модифікації розраховані на роботу з чіткою вхідною інформацією. Проте безліч параметрів у економіці не завжди можна оцінити чітким значенням. Тому в даній статті наведено ще одну модифікацію, яка дає можливість використовувати нечітку вхідну інформацію та не накладає обмежень на вхідні дані.

ПОСТАНОВКА ЗАДАЧІ

Необхідно побудувати модель оцінки тижневого доходу універсаму для визначення впливу зовнішніх факторів на його роботу. Ця модель повинна прогнозувати тижневий дохід універсаму в точці простору, яка описується такими факторами:

- кількість проживаючих у зоні досяжності універсаму за 5 хвилин ходьби, осіб;
- за 10 хвилин ходьби, осіб;
- за 15 хвилин ходьби, осіб;

- кількість конкурентів універсаму, шт.;
- кількість об'єктів притягання (розважальних комплексів тощо), шт.;
- взаємне розташування точок притягання та точок конкурентів у 20-хвилинній зоні досяжності універсаму, хвилин ходьби;
- абсолютне значення об'єктивних ознак привабливості універсаму для покупців;
- час роботи, годин;
- постійність асортименту;
- наявність паркування — максимальна кількість машин, шт.;
- розміщення по відношенню до транспортного потоку;
- наявність вільних площ для маневру покупців;
- якість організації прикасової зони;
- наявність процедур знижок у магазині;
- оцінка доходів потенційних клієнтів універсаму.

Вихідним положенням будемо вважати повну відсутність уяви про структуру моделі та її належності до певного класу. Перелік факторів, наведених вище, можливо, не лише не повний, а й виключає базові параметри. В той же час в перелік можуть бути включені залежні між собою параметри.

Значення (кількісне або якісне) факторів носить оціночний характер — є нечіткою величною. Це може бути інтервал, лінгвістична змінна як впорядкована, так і невпорядкована. Деякі фактори виражаються їх статистичними оцінками (відомі оцінки математичного сподівання та дисперсії для факторів, що мають гауссівський розподіл, або ж параметри будь-якого іншого розподілу). А ще це можуть бути нечіткі числа з різними видами функції належності, представлені параметрами цих розподілів.

ФОРМАЛЬНА ПОСТАНОВКА ЗАДАЧІ

Маємо набір параметрів зовнішніх та внутрішніх характеристик роботи універсаму x_i, $i = 1 \ldots n$ та M вимірів цих параметрів.

Потрібно знайти залежність $n+1$-го параметру від інших.

$$x_{n+1} = F(x_1, \ldots, x_n).$$

ВИБІР МЕТОДУ ПОБУДОВИ МОДЕЛІ

На сьогоднішній день проблеми моделювання складних економічних систем, взагалі, можуть бути вирішені за допомогою дедуктивних логіко-математичних або індуктивних переборних методів. Дедуктивні та імітаційні методи мають переваги у випадку досить простих задач моделювання, коли відома теорія об'єкту, і тому можлива розробка моделі, виходячи з фізичних принципів, застосовуючи знання людини щодо процесів у об'єкті.

Прийняття рішень у таких сферах, як аналіз процесів у фінансовому прогнозуванні, вимагають засобів, здатних отримувати точні моделі на основі прогнозів процесів. Між тим виникають проблеми, пов'язані з великим числом змінних та дуже малою кількістю спостережень і невідомими динамічними зв'язками між змінними. Такі економічні об'єкти є складними слабко обумовленими системами, що характеризуються:

- недостатньою апріорною інформацією;
- великою кількістю параметрів, що не вимірюються;
- зашумленими або короткими вибірками даних;
- слабко обумовленими об'єктами з розмитими характеристиками.

Такі проблеми не можуть бути розв'язані дедуктивними логіко-математичними методами з достатньою точністю. У цьому випадку здобуття знань з даних, тобто знаходження моделі на основі експериментальних вимірів, має переваги у випадку досить складних об'єктів. Такі об'єкти містять мінімальне апріорне знання або не мають визначеної теорії взагалі, що особливо вірно для об'єктів з розмитими характеристиками [2].

Самоорганізація відноситься до емпіричних методів моделювання. Ці методи мають деякі переваги у порівнянні з теоретичними та напівемпіричними методами побудови моделей. В тих випадках, коли спостерігаються лише параметри досліджуваного об'єкту, але не відома структура та механізм взаємодії між елементами складної системи, поведінка якої визначає значення параметрів, підхід самоорганізації виявляється єдиним надійним засобом для побудови моделей прогнозування. За допомогою самоорганізації розв'язок можна визначити, навіть якщо іншими способами отримати результати неможливо. Моделі, отримані за допомогою самоорганізації, мають специфічну сферу застосування і особливо ефективні для прогнозування як на короткі, так і на довгі періоди. Фізичні моделі на основі математичної теорії спостережуваних об'єктів можуть наслідувати лише повністю визначені цілі (ідентифікація та прогноз). Тому побудова моделей у відповідності до нових об'єктивних методів самоорганізації робить можливим замість припущень та грубих помилок запропонувати моделі, які основані на надійній інформації та отримані за допомогою самоорганізації.

Треба звернути увагу на те, що будується модель прогнозу не результату неперервного процесу, а результатів паралельних проявів процесів, що, за нашим припущенням, мають однакову модель. Це одразу відсікає можливість використання для прогнозу методів, які використовують авторегресію.

Зважаючи на викладене вище, поставлена задача може бути вирішена за допомогою модифікованого МГУА з нечіткими входами, який знаходить знання про об'єкт безпосередньо з вибірки даних. Це індуктивний переборний метод самоорганізації з перевагами для досить складних об'єктів, які не мають визначеної теорії, зокрема для об'єктів з розмитими характеристиками. Алгоритми МГУА знаходять єдину оптимальну для кожної вибірки модель за допомогою повного перебору всіх можливих моделей-кандидатів та операції їх оцінки за зовнішнім точним чи балансним критерієм [5, 6] на незалежній підвибірці даних.

МЕТОД ГРУПОВОГО УРАХУВАННЯ АРГУМЕНТІВ З НЕЧІТКИМИ ВХОДАМИ (МГУАНВ)

У відповідності до поставленої задачі модифіковано НМГУА.

Основна ідея

Даний алгоритм, як і НМГУА, використовує поняття лінійної інтервальної моделі

$$Y = A_0 + A_1 z_1 + \ldots + A_m z_m, \tag{1}$$

де A_0 — нечітке число інтервального вигляду, яке описується парою параметрів $A_0 = (a_0, c_0)$; $c_0 > 0$; a_0 — центр інтервалу; c_0 — його ширина; A_i — дійсні числа, але відрізняється тим, що у моделі змінні $z_1 \ldots z_m$ — це нечіткі числа інтервального вигляду, які описуються парою параметрів $z_i = (a_i, b_i)$, $a_i < b_i$, де a_i — ліва границя інтервалу; b_i — права границя.

Тоді Y — нечітке інтервальне число, параметри якого визначаються наступним чином:

ліва границя інтервалу

$$a_y = \sum_{i=1}^{m} A_i a_i + a_0 - \frac{c_0}{2}, \tag{2}$$

права границя

$$b_y = \sum_{i=1}^{m} A_i b_i + a_0 + \frac{c_0}{2}. \tag{3}$$

Для того щоб інтервальна модель (1) була коректною, необхідно, аби справжнє значення залежної величини навчальної вибірки (тобто Y_k, яке також має інтервальний вигляд $\left(a_y^{(k)}, b_y^{(k)}\right)$) належало інтервалу, визначеному формулами (2), (3).

$$\begin{cases} \sum_{i=1}^{m} A_i a_i + a_0 - \frac{c_0}{2} \le a_y^{(k)}, \\ \sum_{i=1}^{m} A_i b_i + a_0 + \frac{c_0}{2} \ge b_y^{(k)}, \end{cases} \quad k = 1 \ldots M. \tag{4}$$

При побудові моделі оптимальної складності розглядалися часткові описи вигляду

$$A_{00} + A_{01} x_1 + A_{02} x_2 + A_{12} x_1 x_2 + A_{11} x_1^2 + A_{22} x_2^2. \tag{5}$$

Через це змінні z_i з формул (1)–(5) записуються таким чином:

$$z_1 = x_1,\ z_2 = x_2,\ z_3 = x_1 x_2,\ z_4 = {x_1}^2,\ z_5 = {x_2}^2, \tag{6}$$

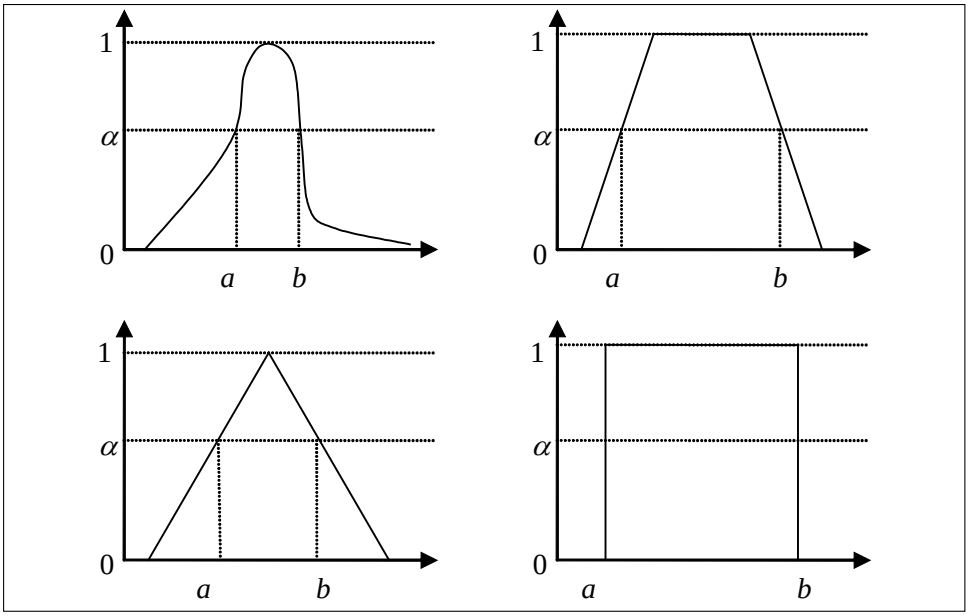
де $x_i x_j = \left(\min(a_i a_j, a_i b_j, a_j b_i, b_i b_j),\ \max(a_i a_j, a_i b_j, a_j b_i, b_i b_j)\right)$.

Нормування вхідних даних

Зведення нечітких вхідних даних до інтервального виду проводиться за допомогою взяття інтервалу α-рівня від функцій належності нечітких величин (див. рисунок).

Дійсні числа замінюються нечітким інтервалом шириною 0. Взяття α-рівня від такого числа завжди буде інтервалом шириною 0.

Нечіткі числа описуються своєю функцією належності та її параметрами. Взяття α-рівня від нечіткого числа буде інтервалом, який визначатиметься точками перетину лінії α-рівня з функцією належності.



Зведення вхідних змінних до інтервального виду

Впорядковані лінгвістичні змінні з N можливими значеннями ($p = 1,2 \ldots N$) замінюються нечіткими числами з трикутною функцією приналежності, яка має параметри A — центр та C — ширина інтервалу. Причому

$$A = p,\ C = 2. \tag{7}$$

Метод оцінювання параметрів лінійної інтервальної моделі

Нехай маємо M вимірів $(n+1)$-ї змінної. Перші n — незалежні вхідні змінні, а $(n+1)$-ша залежить від них, причому характер залежності невідомий. Оцінювання параметрів лінійної інтервальної моделі для часткового опису зводиться до знаходження параметрів A_i, a_0, c_0 таких, що виконуються умови

$$1.\ \begin{cases} \sum_{i=1}^{m} A_i a_i + a_0 - \dfrac{c_0}{2} \le a_y^{(k)}, \\ \sum_{i=1}^{m} A_i b_i + a_0 + \dfrac{c_0}{2} \ge b_y^{(k)}, \end{cases} \quad k = 1 \ldots M, \tag{8}$$

тобто точні значення функціонально залежної величини повинні знаходитися в інтервалі $\left(a_y^{(k)}, b_y^{(k)}\right)$.

2. Ширина інтервалу повинна бути мінімізована.

Такі вимоги зводяться до задачі лінійного програмування

$$\min\left(A_1\sum_{k=1}^{M}(b_1^{(k)} - a_1^{(k)}) + \ldots + A_{p-1}\sum_{k=1}^{M}(b_{p-1}^{(k)} - a_{p-1}^{(k)}) + A_p\sum_{k=1}^{M}(b_p^{(k)} - a_p^{(k)}) + c_0 M \right),$$

$$A_1 a_1^{(k)} + \ldots + A_{p-1}a_{p-1}^{(k)} + A_p a_p^{(k)} + a_0 - \frac{c_0}{2} \le a_y^{(k)},$$

$$A_1 b_1^{(k)} + \ldots + A_{p-1}b_{p-1}^{(k)} + A_p b_p^{(k)} + a_0 + \frac{c_0}{2} \ge b_y^{(k)}, \qquad (9)$$

$$k = \overline{1, M},$$

$$c_0 \ge 0.$$

Наша мета — мінімізувати область зміни вихідних значень Y за рахунок знаходження таких значень параметрів A_i, a_0 та c_0, які б забезпечували мінімальне розсіювання величини Y при одночасному виконанні умови, що справжні значення прогнозованої величини знаходяться у цьому інтервалі.

Оскільки на параметри A_i та a_0 не накладаються умови невід'ємності, то для розв'язання задачі оцінювання параметрів більш зручно перейти до двоїстої задачі лінійного програмування.

$$\max\left(-\sum_{k=1}^{M} a_y^{(k)}\delta_k + \sum_{k=1}^{M} b_y^{(k)} \cdot \delta_{k+M} \right),$$

$$-\sum_{k=1}^{M} a_1^{(k)}\delta_k + \sum_{k=1}^{M} b_1^{(k)}\delta_{k+M} = \sum_{k=1}^{M}(b_1^{(k)} - a_1^{(k)}),$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$-\sum_{k=1}^{M} a_{p-1}^{(k)}\delta_k + \sum_{k=1}^{M} b_{p-1}^{(k)}\delta_{k+M} = \sum_{k=1}^{M}(b_{p-1}^{(k)} - a_{p-1}^{(k)}),$$

$$-\sum_{k=1}^{M} a_p^{(k)}\delta_k + \sum_{k=1}^{M} b_p^{(k)}\delta_{k+M} = \sum_{k=1}^{M}(b_p^{(k)} - a_p^{(k)}), \qquad (10)$$

$$-\sum_{k=1}^{M}\delta_k + \sum_{k=1}^{M}\delta_{k+M} = 0,$$

$$\frac{1}{2}\sum_{k=1}^{M}\delta_k + \frac{1}{2}\sum_{k=1}^{M}\delta_{k+M} \le M,$$

$$\delta_k \ge 0.$$

Ця задача завжди має розв'язок. У відповідності до теореми двоїстості оптимальні рішення прямої та двоїстої задачі співпадають. Тому, розв'язавши двоїсту задачу симплекс-методом та знайшовши оптимальні значення двоїстих змінних, ми зможемо знайти значення змінних A_i, a_0, c_0, $i = \overline{1, C_{r+1}^2}$,тобто визначити параметри лінійної інтервальної моделі.

Опис процедури селекції найкращих часткових описів

Для оцінки якості деякого часткового опису шуканої залежності використовуються такі критерії:

1. *Критерій регулярності за перевірочною вибіркою.*

Вся вибірка наявних даних (довжиною N) ділиться на навчальну та перевірочну: $N_{\text{навч}}$ та $N_{\text{перев}}$. Критерій регулярності визначає суму середньоквадратичних відхилень відповідних границь часткового опису на перевірочній вибірці.

$$\bar{\varepsilon}_i^{\,2} = \frac{1}{N_{\text{перев}}} \sum_{k=1}^{N_{\text{перев}}} \left(\left(a_y^{(k)} - a_{\tilde{y}_i}^{(k)} \right)^2 + \left(b_y^{(k)} - b_{\tilde{y}_i}^{(k)} \right)^2 \right), \qquad (11)$$

де i — номер часткового опису; k — номер перевірочної точки; $a_{\tilde{y}_i}^{(k)}$, $b_{\tilde{y}_i}^{(k)}$ — вихід i-го часткового опису на k-й перевірочній точці.

2. *Критерій незміщеності.*

В основі цього критерію є дуже важливий факт: для одного й того ж об'єкта дослідження за різними вибірками даних та при інших рівних умовах повинні бути отримані близькі моделі, які описують поведінку об'єкта. Критерій має вигляд

$$n_{\text{незм}} = \frac{1}{R_1 + R_2} \sum_{r=1}^{R_1+R_2} \left(\left(a_{\tilde{y}_1}^{(r)} - a_{\tilde{y}_2}^{(r)} \right)^2 + \left(b_{\tilde{y}_1}^{(r)} - b_{\tilde{y}_2}^{(r)} \right)^2 \right), \qquad (12)$$

де R_1, R_2 — розміри навчальної та перевірочної вибірок.

3. *Комбінація критеріїв регулярності та незміщеності.*

Вона являє собою лінійну комбінацію критеріїв (11) та (12), тут α — деякий ваговий коефіцієнт, $0 \le \alpha \le 1$

$$K_{\sum} = \alpha \varepsilon^2 + (1 - \alpha) n_{\text{незм}}. \qquad (13)$$

Загальний опис алгоритму

1. Вибір вигляду часткового опису для залежності, що відшукується.

2. Вибір зовнішніх критеріїв, на основі яких на кожному етапі синтезу буде відбиратися F найкращих моделей.

3. Встановлення лічильника числа моделей k і номера поточного покоління синтезу моделей r в 0.

4. Побудова нового часткового опису. Обчислення критеріїв, $k = k + 1$.

5. Якщо $k = C_F^2$, то $k = 0$, $r = r + 1$. Обчислення середнього значення комбінованого критерію (13) — $N_{\text{см}}^r = \frac{1}{C_F^2} \sum_{i=0}^{C_F^2 - 1} K_i^{\Sigma}$. Якщо $r = 1$, то йдемо на крок 4, інакше — на крок 6.

6. Якщо $\left| N_{\text{см}}^r - N_{\text{см}}^{r-1} \right| \le \varepsilon$, то на крок 7, інакше — обираємо F найкращих моделей у відповідності до зовнішнього критерію (13) та йдемо на крок 4.

7. Вибір найкращої моделі з F оптимальних моделей останнього етапу у відповідності до зовнішнього критерію (13).

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ

Для перевірки та порівняння методів бралися оцінені значення вхідних даних для однієї з мереж київських супермаркетів.

Для порівняння результатів методів НМГУА та МГУАНВ значення нечітких параметрів вхідних даних до методу НМГУА подавалися у вигляді відповідних математичних сподівань, а до МГУАНВ — у нечіткому вигляді.

Порівняльна таблиця методів НМГУА та МГУАНВ за основними характеристиками

Параметри	НМГУА	МГУАНВ
Середньоквадратичне відхилення	6333	4331
R^2	0,40	0,42
Середня ширина інтервалу прогнозу	23788	12208
Середня абсолютна похибка	4913	3569

Як видно з порівняльної таблиці, МГУАНВ на даній задачі та при даних вхідних параметрах продемонстрував кращі значення показників прогнозу на перевірочній вибірці. Середньоквадратичне відхилення від істинних значень майже у 1,5 рази менше, коефіцієнт кореляції кращий на 5%, вдвічі менший інтервал ширини прогнозу, на 20% краще значення середньої абсолютної похибки.

Основним досягненням МГУАНВ у порівнянні з НМГУА є зменшення інтервалу ширини прогнозу. Причиною такого досягнення можна вважати те, що інтервал ширини прогнозу у методах НМГУА та МГУАНВ вираховується за двома різними принципами. У НМГУА інтервал з'являється за рахунок нечітких коефіцієнтів моделі, а у МГУАНВ за рахунок нечітких значень вхідних параметрів. Можна припустити, що саме з цієї причини ми отримали кращу модель.

ВИСНОВКИ

Виходячи з результатів експериментальної побудови моделі, на основі аналізу середньоквадратичного відхилення її похибки та коефіцієнта множинної детермінації R^2, зробимо висновки:

1. МГУАНВ можна застосовувати для побудови адекватних моделей економічних об'єктів. Він дозволяє створити модель об'єкту навіть тоді, коли інформація про його структуру відсутня.

Результати було порівняно з моделлю НМГУА, побудованою на цих же даних.

2. При застосуванні нечіткої вхідної інформації краще МГУАНВ, ніж НМГУА, будує модель економічного об'єкту універсаму. Порівнювались такі параметри: СКВ похибки прогнозу на контрольній виборці, R^2 моделі, середня ширина інтервалу прогнозу.

3. На моделі універсаму МГУАНВ дає значно вужчий інтервал прогнозу, ніж НМГУА.

Отже, при необхідності побудови моделі та прогнозування економічного об'єкту універсаму за наявності нечітких, інтервальних та лінгвістичних вхідних даних слід застосовувати МГУАНВ — метод групового урахування аргументів з нечіткими входами, який було розроблено в даній роботі.

РЕКОМЕНДАЦІЇ

МГУАНВ має дуже широку та перспективну сферу застосування.

Це пов'язано з перевагами нечітких методів внаслідок використання нечіткої інформації, оскільки у більшості випадків навколишнє середовище неможливо описати чіткими однозначними параметрами.

Розроблений програмний продукт GMDHfi можна застосовувати для побудови моделей та прогнозування на основі нечіткої вхідної інформації.

Запропонований метод можна розвивати за такими напрямками: застосування МГУАНВ до моделювання та прогнозування інших макроекономічних, екологічних, медичних та військових об'єктів, а також вдосконалення методології його застосування.

ЛІТЕРАТУРА

1. *Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д.* Принятие решений на основе самоорганизации. — М.: Сов.радио, 1967. — 280 с.
2. *Івахненко О.Г., Івахненко Г.О.* Індуктивні методи прогнозування та аналізу складних економічних систем. — Київ: НІСД, 1997. — 24 с.
3. *Зайченко Ю.П., Кебкал О.Г., Крачковський В.Ф.* Нечіткий метод групового врахування аргументів та його застосування в задачах прогнозування макроекономічних показників // Наук. вісті НТУУ «КПІ». — 2000. — № 2. — С. 18–26.

4. *Зайченко Ю.П, Заєць І.О.* Синтез та адаптація нечітких прогнозуючих моделей на основі методу самоорганізації // Наук. вісті НТУУ «КПІ». — 2001. — № 3. — С. 18–26.
5. *Madala H.R., Ivakhnenko A.G.* Inductive Learning Algorithms for Complex Systems Modeling. — Boca Raton: CRC Press Inc., 1994. — 65 p.
6. *Mueller J.A., Ivakhnenko A.G.* Selbstorganisation von Vorhersagemodellen. — Berlin: VEB Verlag Technik, 1984. — 271 p.
7. *Ivakhnenko A.G., Osipenko V.W.* Algorithms of Transformation of Probability Characteristics into Deterministic Forecast // Sov. J. of Automation and Information Sciences. —1982. — **15**, № 2. — P. 7–15.
8. *Aksenova T.L., Yurachkovsky Yu.P.* A Characterisation at Unbiased Structure and Conditions of Their J-Optimality // Sov. J. of Automation and Information Sciences. — 1988. — **21**, № 4. — P. 36–42.
9. *Beer S.* Cybernetics and Management. — London: English Univ.Pres, 1959. — 280 p.
10. *Belogurov V.P.* A criterion of model suitability for forecasting quantitative processes // Sov. J. of Automation and Information Sciences. — 1990. — **23**, № 3. — P. 21–25.
11. *Gabor D.* Perspectives of Planing. Organisation of Economic Cooperation and Development. — London: Emp. College of Sci. and Technology, 1971. — 347 p.
12. *Ивахненко А.Г., Степанко В.С.* Помехоустойчивость моделирования. — Киев: Наук. думка, 1985. — 216 с.
13. *Ивахненко А.Г., Юрачковский Ю.П.* Моделирование сложных систем по экспериментальным данным. — М.: Радио и связь, 1986. — l18 с.
14. *Stepashko V.S.* Asymptotic Properties of External Criteria for Model Selection // Sov. J. of Automation and Information Sciences. — 1988. — **21**, № 6. — P. 84–92.
15. *Stepashko V.S., Kostenko Ju.V.* GMDH Algorithm for Two-Level Modeling of Multivariate Cyclic Processes // Sov. J. of Automation and Information Sciences. — 1987. — **20**, № 4. — P. 76–84.

Поступила 08.12.2006